1974

Mechanical translation of German into English in the environment of computer assisted instruction /

Mark Alan Bendas
Lehigh University

MECHANICAL TRANSLATION OF GERMAN INTO ENGLISH IN THE ENVIRONMENT OF COMPUTER-ASSISTED INSTRUCTION

by

Mark A. Bendas

A Thesis

Prsented to the Graduate Committee

of Lehigh University

in Candidacy for the Degree of

Master of Science

in

Information Science

Lehigh University

1974

This thesis is accepted and approved in partial fulfullment of the requirements for the degree of Master of Science.

3 May 1974
(date)

Robert J. Baum
Professor in Charge

Robert J. Baum
Chairman of Department

Acknowledgements

I would like to thank Dr. Robert F. Barnes, Head — Division of Information Sciences Lehigh University, for help and guidance in the preparation of this thesis, and Dr, Alexander Waldenrath, Assistant Professor of German Lehigh University, for his help with the German aspects of this paper. My thanks go to M. I. T. Press, William Locke and A. Donald Booth, editors of Machine Translation of Languages, and A. G. Oettinger, author of "The Design of an Automatic Russian-English Technical Dictionary" for permission to reproduce Figures 1-3. Finally, I am grateful to Laura Gottdenker for her help in typing the manuscript.

Table of Contents

Table of Contents
(Continued)

List of Figures

Mechanical Translation of German
into English in the Environment
of Computer-Assisted Instruction

by

Mark A. Bendas

Abstract

The problem of mechanical translation is examined from the earliest work in the early 1950's to work being done currently. The linguistic and programming problems encountered by the early researchers are also discussed.

Two systems are examined in depth. These are the fulcrum approach of Paul Garvin with its emphasis on heuristics, and the use of a mechanical pidgin translator and pidgin as an intermediate language by Margaret Masterman.

This paper investigates pre-editing and post-editing, with the author favoring only limited pre-editing and no post-editing at all. The linguistic problems of selection and arrangement are presented, and the opinion is put forth that any successful mechanical translation algorithm must solve these problems.

The place of foreign language study and mechanical translation in computer-assisted instruction (CAI) is

examined, and a CAI experiment at SUNY — Stony Brook is cited.

The author presents his German-English translation program in the SNOBOL4 programming language, giving sample sentences. Some of the features of the program are its ability to handle various question word orders; a future tense construction; a genetive-possessive construction and its ability to correctly translate sentences containing the ambiguous German pronouns 'sie' and 'ihr'.

The author looks at the possible application his program has in CAI for beginning German students. The paper ends with appendices listing the vocabulary used by his system, the sample sentences, and user and system manuals.

1. An Historical Look at Mechanical Translation

When it became obvious in the 1950's that computers had the capability for language processing applications, linguists and programmers both looked at mechanical translation as an obvious application. This was because translation is a task which requires extensive manipulation of words, such as dictionary look-up, reconstruction of phrases, etc. The computer seemed to be a natural tool to be used in this time-consuming task. However, early in the development of mechanical translation algorithms, it became obvious that the process would be more difficult than at first imagined.

The difficulties were of many types. One was simply a result of the state of the computer hardware and of the nature of the translation process itself.

The translation process basically involves two programming operations. The first is the look-up and matching operation. Here the vocabulary list of the source language is examined, and each entry is matched against the word from the input sentence. The second operation is the algorithm for selecting the proper entry in the target language (if more than one exist) and also for creating the proper word order.

The ratio of look-up to algorithmic programming varied from system to system, but Paul S. Garvin felt in 1958 that as the storage space for the look-up table increased the storage needed for the algorithmic operation would decrease and _vice versa_. (Garvin 1958, p. 78) This would, at first glance, seem not to be the case. One could hold that as one brought new terms into the lexicon of translatable terms one would also bring in cases in which these terms required special treatment. In this case the storage needed for the algorithmic aspects of the translating process would, of course, be increased also. Yet if expanding the storage for look-up means more than simply adding stems , the relation that Garvin hypothesizes holds. If terms including endings (rather than simply stems) or even often-used phrases are included in the look-up, the computation and resulting storage needed to perform the otherwise included procedures of stemming, analyzing and reconstructing phrases, etc. , is eliminated. In 1967 W. J. Plath echoed this point of view and supported Garvin's prediction. In "Multiple-path Analysis and Automatic Translation," which appeared in _Machine Translation_ edited by A. D. Booth, Plath says:

> Whereas in most early MT work the automatic dictionary constituted virtually the entire translation apparatus, the long-term trend has been toward the development of considerably

more complex systems in which the dictionary represents but a single component, albeit an indespensible one. (Plath, p. 298)

Thus, as the algorithmic (non-dictionary) storage increases, as the system becomes more complex, the dictionary is reduced to simply one component of the system. In any case, however, storage requirements led to problems fifteeen or twenty years ago. The machines of the 1950's generally did not have the storage capabilities and rapid access ability needed to store and rapidly retrieve a comprehensive vocabulary in addition to storing the instructions for the translating procedure itself. Garvin discussed this problem in the report cited above.

> It is quite obvious that, in any type of translation program, some kind of bilingual dictionary will have to be stored in the machine memory; in order to allow more than trivial translatability, a dictionary of considerable size will have to be contemplated. It is equally obvious that in any translation program input units will have to be matched one after the other in rapid succession against the units contained in the storage glossary. The rapid access storage requirements of machine translation are far in excess of those required in current mathematical and logical computations; in the latter, extensive storage may be required, but without random access...
>
> Technically then, the problem is not the extent of storage but the requirement that any unit stored in the extensive memory be

> available for immediate look-up. At the present time, extensive storage is possible economically on devices with slow access; rapid-access memory devices are as yet of somewhat limited capacity. (Garvin 1958, pp. 78-79)

Another of the difficulties was one that has generally remained even as the machines themselves have become more powerful. This was a basic linguistic problem. Following the earliest research, it became obvious that translation was more than simply a word-for-word substitution from the source language to the target language. One also had to develop a procedure that took into account the proper word orders of both languages. That is, no unnatural rearranging of the input language's word order should be required to fit the word order of the output language. This task became increasingly difficult, and various compromise systems that required extensive pre-editing and post-editing resulted. Several of these amounted to little more than automatic dictionaries that gave "pseudo-translations which required a human interpreter to put them into readable form. One example of this type of translation system was a Russian-English translator developed by A. G. Oettinger in 1955. (Oettinger, p. 47) In this system the output given by the computer had all possible alternatives for each unit. It also made no attempt to put the translation into the proper word order of the target language. (See Figs. 1-3)

DOKLADY AKADEMII NAUK SSSR
1950, Vol. 70, No. 3

A. G. Lunts

THE APPLICATION OF BOOLEAN MATRIX ALGEBRA TO THE ANALYSIS AND SYNTHESIS OF RELAY CONTACT NETWORKS

Communicated by Academician A. N. Kolmogorov, Nov. 30, 1949.

In recent times Boolean algebra has been successfully employed in the analysis of relay networks of the series-parallel type.(1-3) This algebra is inadequate, however, for a theory of more general networks and for a theory of multi-terminal networks. The purpose of the present paper is to employ a Boolean matrix algebra for investigations of this nature and to describe a series of results obtained along this line.

1. Boolean Matrix Algebra

Let a be a Boolean algebra.(1) We will examine matrices composed of elements from a. As in the case of ordinary matrices (made up of elements from a field) we may define for matrices composed of elements from a the operations of addition and multiplication, which we will write as: $A + B$, $A \times B$. The associative, commutative (for addition) and distributive laws apply in this case.

We introduce the concept of the determinant of a square matrix containing elements from a, as the sum of the

Figure 1. Human translation of a sample Russian text in Oettinger's study.

(Reprinted with permission of the publisher.)

(APPOSITION, Enclosure, Appendix, Application) MATRIX BOOLEAN ALGEBRA (TO, Towards, By, For) ANALYSIS (AND, N) SYNTHESIS RELAY-CONTACT (CIRCUIT, Diagram, Scheme).

(IN, At, Into, To, For, On, N) (LAST, latter, new, latest, lowest, worst) (TIME, tense) FOR ANALYSIS (AND, N) SYNTHESIS RELAY-CONTACT ELECTRICAL (CIRCUIT, diagram, scheme) PARALLEL- (SERIES, successive, consecutive, consistent) (CONNECTION, junction, combination) (WITH, from) (SUCCESS, luck) (TO BE UTILIZE, to be take advantage of) APPARATUS BOOLEAN ALGEBRA. BUT THIS APPARATUS (TO FIND X-SELF, to turn out, to be found, to prove) (INSUFFICIENT, inadequate, scanty) FOR THEORY (CIRCUIT, diagram, scheme) (GENERAL, common) TYPE, (BUT, and, yet, if, while) ALSO FOR THEORY MULTIPOLAR (CIRCUIT, diagram, scheme). (IN, At, Into, To, For, On, N) (PRESENT, genuine) (ARTICLE, item, clause) (TO BE OFFER, to be propose, to be suggest) FOR (INVESTIGATION, research, analysis, exploration, paper, essay) (SUCH, so, a sort of) (SORT, kind, family, genus, gender) (TO UTILIZE, to take advantage of) MATRIX BOOLEAN ALGEBRA (AND, N) TO BE DESCRIBE (ROW, series) RESULT, GOTTEN (IN, at, into, to, for, on, N) THIS (DIRECTION, trend, order, permit).

I. MATRIX BOOLEAN ALGEBRA

(LET, Though) *a* (TO BE, to eat, O.K.) SOME BOOLEAN ALGEBRA. TO BE (TO CONSIDER, to examine, to discuss) MATRIX

Figure 2. Output of the same Russian text translated by the automatic dictionary in Oettinger's study.

(Reprinted with permission of the publisher.)

Doklady Akademii Nauk SSSR

1950, Tom LXX, No. 3

A. G. Lunts

(~~Apposition, Enclosure, Appendix,~~ Application) of Matrix Boolean Algebra (To, ~~Towards, By, For~~) the Analysis (And, ~~N~~) of Synthesis Relay-Contact (Circuit$ ~~Diagram, Scheme~~).

(In, ~~At, Into, To, For, On, N~~) ~~(last, latter,~~ ~~new,~~ recent ~~latest, lowest, worst~~) (time$ ~~tense~~) the for analysis (and, ~~N~~) synthesis of relay-contact$ electrical (circuit$ ~~diagram, scheme~~) containing parallel-(series, ~~successive, consecutive, consistent~~) (~~connection, junction,~~ combination) of (with, ~~from~~) (success, ~~luck~~) (~~to be~~ has been utilized, ~~to be taken advantage of~~) the apparatus of Boolean algebra. But this apparatus (~~to find a self, to turn out, to be found, to~~ proved) (~~insufficient,~~ inadequate, ~~scanty~~) for the theory of (circuit$ ~~diagram, scheme~~) of general, ~~common~~) type, (~~but~~, and, ~~yet, if, while~~) also for the theory of multipolar? probably O.K. (circuit$ ~~diagram, scheme~~). ~~(In, At, Into, To, For, On, N~~) The (present, ~~genuine~~) (article,

Figure 3. Sample work sheet of a human post-editor for the computer generated output text in Oettinger's study.

(Reprinted with permission of the publisher.)

As one can see from the above figures, extensive post-editing was needed. This process resembled many of the attempts at machine translation which found the needed algorithms too difficult to formalize. This problem continued to plague attempts to develop machine translation procedures throughout the late fifties and early sixties.

As the difficulties involved in machine translation became more apparent, much of the work that was being done ceased. This was due not only to the difficulty of the task, but to the realization that other language processing tasks (e.g. abstracting, indexing, document retrieval, etc.) were more feasable than ever before because of faster and larger computers, and were much less complex than machine translation. Thus, in the mid- and late sixties, work in machine translation slowed so that only a few specialized projects remained.

One must then ask why machine translation was not helped by the same advances in computers that aided other forms of automatic language processing. The answer to this is twofold. First, the new larger and faster computers did in fact help machine translation insofar as the storage capabilities increased, thus allowing for larger vocabularies. The speed increase also helped execute the translation algorithms themselves much faster. Yet the difficulties remained, for the look-up is only a part of the translation process, and for the speed of the computers to help the algorithms themselves, one had to have successful (although perhaps long and involved) algorithms with which to begin.

Consequently, by the latter part of the 1960's, the state of machine translation had not greatly improved. Small advances, of course, had been made, but the broad picture remained relatively unchanged. Jagjit Singh discusses this in his book Great Ideas in Information Theory, Language and Cybernetics, written in 1966.

> The afore-mentioned threefold division of the machine-translation task — word recognition, text translation, and text transposition —[1] represents by and large the grand strategy underlying the present practice of the art. Even though none of these problems, as we have remarked, has been fully solved, the partial solutions so far suggested can actually be implemented on a computer. (Singh, pp. 287-288)

The state of the machine translation studies in the late sixties has also been presented by Harry Josselson in his 1971 overview, "Automatic Translation of Language Since 1960: A Linguist's View," which appeared in Advances in Computers volume 11. He reiterates the linguistic

---

[1] This threefold division is really equivalent to the twofold division presented earlier, with word recognition and translation combined into the look-up.

problems mentioned above and puts the total question quite succinctly, saying:

> ...the primary reason for the unattainability of fully automatic high-quality translation is the very nature of language itself. This fact escaped many early researches in machine translation, particularly those who were hardware-oriented and who proceeded from the very naive position that since everybody speaks a language, everybody should, theoretically at least, be able to deal with any processes which involve language. (Josselson, p. 2)

Perhaps he is speaking of Warren Weaver, whose famous statement Singh supports. Weaver said to Norbert Wiener in a letter to the latter, "When I look at an article in Russian, I say: 'This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.'" (Weaver, p. 18) This, of course, totally ignores the linguistic problems that translation presents.

As Josselson's point of view is that of a linguist, he places the major importance on the linguistic aspects of machine translation. It is thus of little surprise that the projects he cites are for the most part largely theoretically oriented.

As previously stressed, work in machine translation did not bear out the original expectations for success. The early optimism could very well have been the result of misinterpretation of early results. Josselson has this to say.

> However, the first experiments in machine translation, starting with the trial run in 1953 by Leon Dostert, Director of Georgetown University Institute of Languages and Linguistics, really did not prove anything, particularly as far as the possibility of fully automatic high-quality translation is concerned. This was due to the fact that the first MT experiments were carried out on very limited small texts, with bilingual glossaries and grammars specially tailored for these texts, thus in effect creating an ideal, closed linguistic system in contrast to the openness and dynamics of natural language. The computer programs specifically designed for these small texts, of course, guaranteed the success of these experiments. As a consequence early MT researchers arrived at the conclusion that all that was necessary to achieve practical results in MT was to increase the size of the dictionary and to expand the grammar. (Josselson, p. 7)

Consequently research in MT expanded greatly in the early sixties. Yet even as Josselson optimistically reports the existence of over seventy projects in machine translation worldwide in 1969, (Josselson, p. 26) many of the projects that had been started in the early sixties had either ended their research with only minor success or shifted their emphasis to different (though related) fields such as syntactic analysis. (Josselson, p. 26-40)

Thus, at the end of the sixties and until the present the state of machine translation appeared to be at a standstill. There are two different points of view existing now. There are those who still feel that linguistic theory must develop further and so are working on the formalization of linguistic notions. There are others who feel that developing a complete formal description of a language will be almost impossible. Consequently, those with this point of view look at translation heuristically. This paper will examine two systems with different theoretical points of view, Paul Garvin's as expressed in "Machine Translation Today: the Fulcrum Approach and Heuristics" and Margaret Masterman's in "Mechanical Pidgin Translation". The radical difference in these two approaches to the problem of machine translation show the widespread lack of agreement in this matter.

2. Two Different Approaches: Pidgin Translation and the Fulcrum Approach

In her article Masterman discusses work done throughout the sixties on various pidgin translation projects including her own. Her system does not even attempt to supply a complete translation of an input, but gives only the "pidgin English" intermediate translation which is then interpreted by the person interested in the translation. Pidgin English is a strange hybrid language with the vocabulary of English words and a grammatical structure similar to Chinese. The characteristics of a pidgin dictionary are as follows:

(1) Predominance of dictionary entries for phrases rather than words.
(2) Special sub-dictionaries and the presupposition that a choice of sub-dictionary appropriate to the text has been made.
(3) Specially constructed symbols here called 'pidgin-variables' i. e. widely ambiguous words which the reader intuitively interprets according to the context. (One example is 'che' in Italian, which in pidgin is either 'that' or 'what'.)
(4) The omission of grammatical and syntactic features of the input language that a word-for-word Machine Translation program cannot transform. (Masterman, p. 298)

Mastermen herself defines a Mechanical Pidgin Dictionary with additional requirements. (All italics are hers.)

(5) It must not allow for any alternatives being included in the output between which the reader of the output must find a way to choose.
(6) The program must contain no provision for changing the word-order of the text.
(7) The pidgin must be treated and studied as a homogeneous language with properties of its own, *without consideration of the fact that different specimens of it may be derived from different source languages*. (Masterman, p. 298)

Masterman examines pidgin translation through a study of the output of the system originally designed by Booth and Richens when they first used pidgin as an intermediate language for translating various languages into English. These outputs are then refined by additional procedures designed by Masterman. The outputs also contain special pidgin symbols that are included in the pidgin output. The symbols originally used by Richens and Booth, as well as by Masterman are as follows:

a	accusative	o	oblique
d	dative	p	past
f	future	q	passive
g	genitive	r	partitive
i	indicative	s	subjunctive
l	locative	u	untranslatable
m	multiple, plural or dual	v	vacuous
n	nominative	z	unspecific

(Masterman, p.200)

It would be best to analyze Masterman's system by examining a sample sentence originally translated into pidgin English by Booth and Richens. The sentence is in Italian, and the output in pidgin is given as well as the actual English translation. The asterisks are positioned to denote each phrase to be decomposed individually. The sample Italian sentence is:

> E stato prov * ato che i cereal * i d'invern * o cresc * iuti in serra mosti * ano poc * a resistenza al freddo, mentre gli stessi cresc * iuti in campo apert * o, sono molt * o piu resistent * i. (Booth and Richens, p.37)

The pidgin output is:

is been/status prove p that/which? v cereal m of winter z grow pm in mountain/crowd/ greenhouse show m little v resistance to cold while v same m/is ps grown pm in field open v are much v more resistant m.
(Booth and Richens, p.38)

The "slashes" separate alternatives that would have to be chosen by the reader. This "translation" is at best unclear and may in fact be misleading. One can only get the barest of ideas of what this sentence means. This is how the sentence translates into English.

> It has been proved that winter cereals grown under glass show little resistance to cold, while those grown in the open are much more resistant. (Booth and Richens, p.38)

The first step in the refinement of the sentence was the removal of the z and v entries. The z entry is concerned with an Italian ending which is grammatically ambiguous, while the v entries are for Italian entries that are so ambiguous that they mean nothing at all. Masterman lists the z and v dictionary entries that produce the output.

ITALIAN	
i	v
-o	z
-a	v
-e	v

(Masterman,p.202)

Masterman holds that nothing can be done with these, pidginwise, so she deletes them from the output. The resulting sentence is:

> is been/status prove p that/which cereal m of winter grow pm in mountain/crowd/ greenhouse show m little resistance to cold while same m/is ps grown pm in field open are much more resistant m. (Masterman, p.203)

The next step in refining the outputs is to convert m, g, o, p. Masterman describes the method as follows:

> In the Italian case —at, translated by Richens p, we pidginize as '-ed', -i, which he translated by m, a '-s' (on the assumption that the Plant Genetics pidgin-dictionary is never going to have any imperatives...) and '-ano' we pidginize as 'they'.
> The result is as follows:
> is been/status prove-ed that/which cereals of winter grow-ed-s in mountains/crowd/ greenhouse show-they little resistance to cold while same-s/is grow-ed in field open are much more resistant-s. (Masterman, p.204)

The next step is to create any needed pidgin variables. One needed for Italian is '(w)that' for che. (Masterman, p.204) The result following this step is:

> is been/status prove-ed (w)that cereal-s of winter grow-ed in mountain/crowd/ greenhouse show-they little resistance to cold while same-s/is grow-ed in field open are much more resistant-s. (Masterman, p.204)

The final step in the refining of the translation is the exploitation of a very large phrase dictionary (500,000 entries). This equates phrases in the source language with sequences of pidgin variables. The following phrases occur in the Italian sentence.

> <u>e</u> <u>stato</u> — 'has + been'
> <u>cresciuti</u> <u>in</u> <u>serra</u> — 'grown + under + glass'
> <u>gli</u> <u>stessi</u> — 'the + same'
> (Masterman, p.205)

This gives the following final refinement of the original pidgin translation.

> HAS + BEEN PROVED-ED (W)THAT CEREAL-S OF WINTER GROWN + UNDER + GLASS SHOW-ED-THEY LITTLE RESISTANCE TO COLD WHILE THE + SAME GROW-ED IN FIELD OPEN ARE MUCH MORE RESISTANT-S. (Masterman, p.205)

This is much closer to the actual translation shown above than either the pidgin translation or any of its refinements. Masterman comments that the above final refinement was read and understood quite easily by those who were shown it.

Masterman concludes the article with an evaluation of the work already done and courses of action she feels should be taken in this field of machine translation.

> The way forward is::
> (1) To accept the conclusion derivable from The Mechanical Pidgin Translation experiments that the phrase and not the word is the semantic unit of translation;
> (2) To make the machine cut the source text up into phrases (using syntactically and/or phonetically derived data), and then to do a dictionary match of these with a Mechanical Pidgin phrase dictionary in which classes of phrases are coded into sequences of pidgin-variables...

(3) To assign to these sequences of pidgin-variables a mathematically determinate recursive structure which can also be interpreted semantically as a Mechanical Pidgin structure. Thus the notion of a Mechanical Pidgin variable is abstracted from that of an English pidgin-variable; and the notion of the structure of a Mechanical Pidgin from that of a simplified English grammar and syntax;

(4) To print, as first output, the structured concatenation of sequences of pidgin-variables: each such sequence conveying a 'bit of information'. This will be the message;

(5) To convert this output by some phrase-construction program into a sequence of phrases in the target-language. (Masterman, pp. 224-225)

On examining the points of view expressed by Masterman and the others working with Mechanical Pidgin Translation, the shortcomings of mechanical translation become obvious. Masterman (and the others working on the pidgin translation) have given up the goal of total translation, for they rely too heavily on post-editing. Also, even if the various refinements of the pidgin output are performed mechanically, the final output, though understandable, is not really a high quality translation. There is also extensive human intervention throughout, so the use of Mechanical Pidgin Translation as even a stepping-stone to a totally mechanical system is futile.

In her closing comments noted above, Masterman does bring up a critical point. What is the proper semantic unit of translation? She contends that it is the phrase. Practically, however, when dealing with a large vocabulary and wide corpus of texts to be translated, the total number of phrases needed to accomplish translation is too large. For a very specialized and closed system, such as that subject to pidgin translation, however, the use of the phrase is desirable. Yet despite the apparent impracticality for use in a large system, the phrase (or some other unit larger than a word, yet smaller than a sentence) can still have a utility in the development of a machine translation system. This can be seen in the fulcrum approach as expressed by Paul Garvin.

Garvin's article, "Machine Translation Today: the Fulcrum Approach and Heuristics", appeared first in 1968 in Lingua. In it Garvin introduced two main ideas, the use of heuristics in syntactic analysis and translation, and the theory of the fulcrum and its application in machine translation.

It should first be understood what Garvin means by the "fulcrum".

> The concept of the fulcrum implies the use of key elements within the sentence (fulcra) as starting points for the searches performed by the algorithm. This means that the algorithm, in searching through the sentence,

> does not simply progress from word to word, but in fact 'skips' from fulcrum to fulcrum. It performs a little search sequence each time it has reached a fulcrum, and goes on to the next fulcrum when the particular search is completed. (Garvin 1968, pp. 97-98)

From this Garvin presents how this approach differs from other approaches of syntactic analysis and automatic translation.

> 1. The Fulcrum approach favors a bipartite, rather than a tripartite, organization of the parsing system.
> 2. The Fulcrum approach is characterized by two basic operational principles: (a) the concept of the fulcrum; (b) the pass method.
> 3. The Fulcrum approach aims at producing a single interpretation of each individual sentence, rather than producing all conceivable interpretations. (Garvin 1968, p. 95)

The basic difference between the bipartite and tripartite organization is that the tripartite organization has a separate component for the information used by the processor, while the bipartite organization incorporates the information right into the processing algorithm. Thus, Garvin's bipartite organization contains a dictionary component with grammar codes (e. g. noun, verb, etc.) for each entry, and an algorithm which contains processing subroutines as well as the information to do the processing. The tripartite organization contains the dictionary component, the processing algorithm and a separate component for the information the processor uses.

Garvin questions the advantages usually cited by supporters of a tripartite system. The first advantage is that such a system separates the labor of the programmer (designing and maintaining the processor) from that of the linguist (designing and maintaining the table of rules). This eliminates the problem of communication between the two specialists that sometimes exists. The second advantage is that the same processor can be used for different tables of rules, thus giving the linguist much freedom of experimentation. Garvin, however, holds that:

> These advantages apply particularly well to small experimental systems oriented towards linguistic research; for large-scale experimentation, oriented towards the processing of randomly chosen bodies of text with the ultimate aim of designing an operational translation system the advantages of a tripartite system are less clearcut. This is why the Fulcrum approach favors a bipartite organization of the parsing system. (Garvin 1968, p. 96)

Continuing the above discussion, one now looks at Garvin's concept of "multiple passes" through each sentence for the automatic parsing of the sentence. This technique means that more than one "pass" is made through each sentence, with each pass designed to identify a specific set of grammatical conditions required in the process of recognition. Thus, each pass has its own set of fulcra and its own search sequence. (Garvin 1968, p. 98)

Garvin remarks next that only one interpretation is given for each sentence. This is in contrast to more theory-oriented systems (such as Masterman's) which supply all possible parsings. Because the fulcrum approach requires a correct parsing as a step toward mechanical translation, supplying all possible parsings is not only unneeded, but undesirable. The system only considers one correct parsing, regardless of ambiguities. Thus, as soon as a correct parsing is found, the system looks no further.

The final topic Garvin discusses is the place of heuristics in linguistic analysis and machine translation. In the fulcrum approach the heuristic aspect is, in fact, imbedded in the algorithm complete with its own routines. "Thus, the executive routines of the heuristic, which carries out the 'guessing'strategy by calling the trial and evaluation routines, in fact constitutes a bridge between the deterministic main portion of the algorithm and the heuristic portion." (Garvin 1968, p.105)

Garvin's own summary best expresses the heuristic portion of his fulcrum approach.

> 1. The heuristic portion of the Fulcrum algorithm is called whenever there is a possibility that a given identification made on the basis of the immediate context may have to be revised

on the basis of information provided by the broader context.

2. The conditions requiring the use of heuristics are recognized by the deterministic portion of the Fulcrum algorithm.

3. The mechanism for calling the heuristic syntax consists in the writing of a record (setting a 'flag') in the sentence image which the program produces, indicating that a given identification has been made on a trial basis and is subject to heuristic revision.

4. The evaluation criteria for the revision of trial identifications consist in various conditions of manditoriness of occurrence of certain syntactic components...Some of these are implicit others explicit.

5. The mechanism for applying a heuristic revision to a trial identification consists of the following:

(a) The program first notes the absence of a mandatory syntactic element...

(b) The program now tests for the presence of heuristic decision records ('flags') in the sentence image and checks whether the recorded element is a suitable candidate for the missing syntactic component.

(c) If the tests are positive, the trial identification is revised and a definitive identification is substituted for it.
(Garvin 1968, pp. 111-112)

Thus, Garvin's system is a theoretical one in that he himself does not design and implement a mechanical translation system, but his is a system designed to help implement a complete automatic translation of natural language, and many of his ideas have immediate applicability in this area.

The above is a brief general review of machine translation through the sixties with a close examination of two quite different approaches to the machine translation problem. Several concepts of machine translation should, however, be analyzed individually. These will be discussed both generally and as they apply to the German-English translation system developed here.

3. Pre-editing and Post-editing

The obvious goal for most doing research in mechanical translation is complete automatic translation with no human intervention. However, total automatic translation of natural language text has proven to be very difficult, so compromises have had to be made. One of these compromises has been the addition of a pre-editor, a post-editor, or even in some cases both. Pre-editing is basically putting the input text into a form to facilitate translation beyond simply preparing it to be read by the computer. Post-editing, on the other hand, is taking the output from the translation process and having a human edit it for presentation to the user.

Pre-editing and post-editing both result from the assumption, belief or realization that total automatic translation is not feasible. The extent of the pre-editing or post-editing needed depends greatly on the nature of the text and the nature of the user group for whom the translating is being done.

One example, for instance, involves work being done on an English-to-French translator by researchers in Montreal, Canada in conjunction with the officially bilingual Canadian national government. Kathleen Booth discussed that work in "Machine Aided Translation with a Post-Editor". As noted above, the use of a post-editor here is dependent on the nature of the translation being done and on the user group involved.

> It should be remarked, however, that as far as Mechanical Translation is concerned, the object is to produce a rough translation which can be polished by someone having a good knowledge of French, but not necessarily of English. Thus, the essential requirement is to preserve the sense, infelicities of style being removed later. (K. Booth, p. 53)

The use of a post-editor here adds no extra personnel, since translators are in any case needed to translate official Canadian documents. Consequently, use of a post-editor here is not at all undesirable and is appropriate to the task and circumstances.

Originally a pre-editor was thought to be less undesirable than a post-editor. Back in 1955 Erwin Reifler, in "The Mechanical Determination of Meaning", suggested:

> The determination of the intended nongrammatical meaning could be assigned to a pre-editor. He would have to be familiar with the langauge of the original text and, adding

> symbols such as diacritic marks, would increase the semantic explicitness of the conventional graphic form of the original text sufficiently to allow a mechanical system to supply the correct output equivalent in every case. In support of the pre-editor proposal, I pointed out that the task of the pre-editor would be much easier to accomplish than that of a post-editor because the former would determine intended meanings in a context completely intelligible to him...
>
> Thus I concluded at that time that the simplest form of MT would be one with pre-editing, in which the pre-editor determines the meaning indicated by each context and denotes it by special graphic symbols which he adds to the conventional written form in all instances involving multiple meaning. (Reifler, p. 141)

An argument against this reasoning can be made from a purely practical point of view. A good human translator can eliminate the need to read through the text in the object language to get rid of possible semantic ambiguities by use of additional symbols. This is so even if the human pre-editor is not fluent in both languages, because the use of this type of pre-editing assumes high quality translation as its output. If the output can be assumed to be readable, then perhaps the combination of pre-editor and automatic translation system may be less expensive and faster than human translators, although this is not clear and depends heavily on the cost of the system itself.

Limited pre-editing does seem to have its place in certain applications. In a system to be used as a teaching aid, pre-editing by the student can be helpful to him not only to prepare the text for the automatic translating, but also as a learning experience for the student. This is the case in the German-to-English translation system designed here and discussed later. This system will be examined later at great length. However, the only pre-editing in it is for the user to type a plus sign (+) following the last word of the subject of each sentence. In addition to its being necessary for good translation it is obviously an aid for beginning students to understand the sentence structure in the language.

Many researchers shied away from the use of either a pre-editor or a post-editor, despite the belief by some that pre-editing and post-editing were desirable. Leon Dostert discussed this as long ago as 1955 while examining the Georgetown-I.B.M. experiment.

> Two more assumptions oriented the research for the first experiment, as well as what has been done since. The third was to eliminate the idea of post-editing. We set out to feed in the normal language at the input, without prior human processing, and we aimed at obtaining clear, complete statements in intelligible language at the output. This last is not to say that certain <u>stylistic</u> revisions may not be required as we progress, just as when the translation is done by human beings.
> (Dostert, p. 127)

The use of either pre-editing or post-editing depends on the structure and goals of the system. If totally automatic translation is desired, pre- and post-editing would not be included. If, however, economics is not a problem or if the system is specialized so that pre- or post-editing has another use in addition to text revision, then including a pre-editor or post-editor in the system would not be undesirable.

4. Two Linguistic Aspects: Selection and Arrangement

The two major aspects that must be taken care of by every machine translation system are <u>selection</u> and <u>arrangement</u>. <u>Selection</u> deals with that part of the translation procedure that selects the correct term from the dictionary corresponding to the term to be translated in the sentence. It also chooses between cases if a dictionary term has more than one translation. <u>Arrangement</u> is simply putting the translated terms into the proper syntactic word order in the target langauge. If, for example, one is translating a German question into English, one must revise the inverted word order of German questions into the proper word order of English questions. Thus, to have high quality translation without resorting to a post-editor, the arrangement problem must be solved or the resulting translation will be difficult to decipher, or at least awkward.

Because selection and arrangement are basic to any translating system, the necessity for establishing selection and arrangement procedures was recognized early in the research in machine translation. Garvin realized the importance of solving the problems of selection and arrangement in most of his early writings on machine translation. He discussed the problems in depth in "Some Linguistic Problems in Machine Translation: an Early View Still Held." He stated the problem in this way:

> The selection problem can be rephrased as follows; what portion of the total system-derived meaning of the source unit applies in any given textual fragment, and to which of the possible equivalent portions of target units does it correspond? (Garvin 1956, p.69)

Thus, each unit has two meanings in each use. First, it has a broad system-derived meaning, its meaning within the system. Second, it has its narrow individual meaning in the context of the one individual sentence. The role of the selection procedure in this approach is to decide which individual meaning the term must receive from the broad system-derived meaning. It may be best to look at this procedure schematically.

Thus, schematically this concept becomes:

SPECIFIC MEANING (in the object language sentence) to

SYSTEM-DERIVED MEANING (in the object language) to

SYSTEM-DERIVED MEANING (in the target language) to

SPECIFIC MEANING (in the target language translation).

Garvin sums up the selection problem as follows:

> Selection decisions are thus, in terms of the above, contextually determined and situationally determined, and the objective of translation analysis becomes one of singling out the specific determining factors for each decision in the given context and situation. (Garvin 1956, p. 70)

Arrangement decisions, on the other hand, concern not the equivalence of terms, but the equivalence of sequential relations. The objective of any arrangement decision in the translation system is to obtain a sequence in the output language that has the same meaning as the corresponding sentence in the input language, although the order of the terms may not be the same in both cases. To decide whether or not the word order should be changed, the various transformational order relations between the input language have to be established by the translation algorithm. Thus, to decide whether a German question's word order must be changed to give a good translation of it in English, the translation algorithm must have access to the relationship between proper word order for German questions and proper word order for English questions.

Garvin explains that although there are similarites between the processes involved in making a selection decision and those involved in making an arrangement decision, the processes are essentially different.

> We are thus again dealing with problems of meaning equivalence, but instead of the meanings of units we deal with the meanings of relationships, and instead of lexical meaning we are dealing with grammatical meaning, that is, meaning not unique to each given unit but shared by more than one unit. While selection decisions are thus unique and specific to each unit, arrangement decisions are recurrent for classes of units... The above means that, unlike selection decisions, arrangement decisions can — at least as far as the purely linguistic factors are concerned — be considered entirely within the bounds of machine translatability, since the limits of the linguistic context need not be excluded. (Garvin 1956, p. 72)

So, for fully automatic translation, these two problems of selection and arrangement must be solved (without either pre-editors or post-editors) to give readable translation. The extent to which the questions of selection and arrangement are dealt with in the system designed here will be discussed in the later description of the German-to-English translation system.

5. Mechanical Translation and Computer-Assisted Instruction

The next topic is the place of mechanical translation in computer-aided instruction. Is teaching a foreign language on an elementary level a realistic candidate as a process to which computer-aided instruction can be applied?

To answer the above question, one must answer a more basic question; that is, what aspects of foreign language instruction can be computerized. The answer seems to be those types of aspects that are generally acceptable as material for any type of computer-aided instruction, namely subjects about which questions can be formed that have specific answers that the computer can either recognize or reproduce (depending on the structure of the program); those portions of the subject matter that are tedious or time-consuming for the instructor to correct; or exercises that require much drilling for the student. Thus, the translation of sentences is a good candidate for computer-aided instruction, because the translation of sentences requires much practice and drilling by the student; the correction of these translations is time-consuming for the instructor; and at least on the elementary level, the sentences generally have only one or two possible translations, and it is not too difficult for a student to recognize a translation different from his own.

The practicality of computer-aided instruction for foreign language teaching (including sentence translation) was demonstrated by the use of computers to help teach elementary German at the State University of New York at Stony Brook during the academic year 1966-1967. H. W. Morrison and E. N. Adams report the important results of this study in "Pilot Study of a CAI Laboratory in German", which appeared in Computer-Assisted Instruction edited by Richard C. Atkinson and H. A. Wilson. Morrison and Adams describe how the system was established.

> Students in the CAI section met their instructor for three 50-minute class periods each week where they were taught by the direct method including audio-lingual pattern drills. There was no written homework and practically no class time was spent on writing, translating, spelling, vocabulary, or reading. Recitation which emphasized facility in writing German was scheduled for two 50-minute periods each week at a CAI instructional station (an IBM 1050 with auxiliary tape recorder and slide projector). Students could schedule additional time if terminals were available. (Morrison and Adams, pp. 191-192)

The students were encouraged not to proceed to the next lesson until they had mastered the lesson on which they were working. They were told that their grades would not depend on their performance in CAI work.

Adams and Morrison used another section of the class for comparison purposes. This other section had the same instructor, used the audio-lingual method (ALM), and had three class periods and two 25-minute conventional language labs.

Morrison and Adams relate many comparisons between the CAI section and other sections. Perhaps, however, the most effective comparison is that of final grades in the course, where the CAI section showed higher grades than any other section, even though both the CAI section and the instructor's ALM section were both rostered randomly. The results are given below in Figures 4 and 5.

	W/I	F	D	C	B	A
Stony Brook ALM students outside study	12%	10%	12%	25%	28%	13%(n=226)
1st semester ALM section	4%	22%	13%	30%	22%	9%(n=23)
2nd semester ALM section	—	—	6%	29%	35%	29%(n=17)
1st semester ALM students who registered for 2nd sem.	—	—	7%	43%	36%	14%(n=14)
CAI section	4%	0%	8%	38%	35%	15%(n=26)

Figure 4. First semester grades of CAI study. (Morrison and Adams, p. 195)

	W/I	F	D	C	B	A
ALM students outside study	14%	3%	8%	31%	26%	18%(n=163)
2nd semester ALM section	6%	0%	24%	35%	24%	12%(n=17)
1st semester ALM students who registered for 2nd sem.	14%	7%	14%	36%	21%	7%(n=14)
CAI students who registered in CAI for 2nd semester	17%	0%	17%	33%	21%	12%(n=24)

Figure 5. Second semester grades for CAI study. (Morrison and Adams, p. 195)

The inportant lines in Figure 4 above are first, second and fifth. Looking at these one sees that the CAI section had higher percentages of A's, B's and C's than either ALM students in general or the ALM section used in the study. Also for the first semester both ALM groups had at least the same, and in most cases, higher percentages of D's, F's and Withdrawals/Incompletes.than the CAI section. The second semester grades are less conclusive (see Figure 5) but the lack of F's in the CAI section shows some sort of positive comprehension by the students there.

It should be noted that the system described here deals with more than simply mechanical translation. Yet mechanical translation seems to be dealt with in at least a limited fashion. Perhaps the key to success is simply that working at the computer terminal gives the student an added opportunity to practice in a field that requires much tedious, drill-like study in the elementary stages. If a mechanical translation procedure can be developed that works well on elementary sentences with little pre-editing and no post-editing, then it would be useful in instructing beginning students. This is the basis of the translation procedure developed here.

6. Linguistic and Programming Aspects of the German-English Mechanical Translator

The German-English mechanical translator was first designed on a small scale in the WIZARD programming language, but the limitations of formats in WIZARD forced conversion to another, more flexible language, SNOBOL4. A conversion to SNOBOL4 of the code already written in WIZARD was the first step. Addition of new linguistic features and expansion of the vocabulary completed the programming. All programming was done on Lehigh University's CDC 6400 computer.

The goal of the translating program is to give correct English translations of relatively simple German sentences as an aid in the instruction of first year college German students. It would thus give the student the opportunity to practice concepts that were proving difficult for him.

The system is thus able to translate into good, syntactically correct English, simple present tense German sentences as well as those in the compound future tense. It should be noted here that the simple past tense can be handled easily by simply expanding the vocabulary to include the past tense verb forms. Some of the other features include the handling of the ambiguous German pronouns 'sie' and 'ihr', the ability to handle all forms of German question word order and translating it into a correct English question, and the ability to translate genetive-possessive constructions. These features will be discussed later while looking at specific sample sentences.

The way the program normally translates a sentence is as follows. First, the user types in a German sentence (either on a teletype or on a computer card). Here he must perform one act of pre-editing: he must type a plus sign (+) following the last word of the subject of the sentence. This is needed to discover where the subject and the predicate are in the sentence. This bit of pre-editing, in addition to being necessary for correct translation, also helps the student understand the structure of the sentence. Following this input stage, the program then pares off each word and places each in an array, noting the location of the plus sign. The sentence is also searched for a verb whose location is noted. The program then notes whether the sentence is a question or a declarative sentence by looking for a period or a question mark (in this case a '*'). If the sentence is declarative, the translation begins normally; while if it is a question, this fact is noted and the proper transformations are begun. When a word is encountered that is to receive special treatment, the program does so. Otherwise, it is a simple look-up in the dictionary and substitution of the proper English term. If a word is encountered that does not appear in the German lexicon, an appropriate error is printed. This occurs until the entire sentence is translated, and the translation is printed. The process is repeated for each sentence input until 'XXX' is encountered, which indicates the end of the input.

The following is an examination of some sample sentences translated by the German-English mechanical translation program. A complete listing of the vocabulary appears in Appendix,I. The sentences, as they are input and as the system revises them, are included in Appendix II, along with their translations.

(1) ICH HABE EINE SCHWESTER.
I HAVE A SISTER.

This sentence is translated almost without any processing beyond the look-up. Here, however, the selection problem arises. The German indefinite article 'EINE' can be translated as either 'A' or 'AN', depending on whether the first letter of the following word is a consonant or a vowel. When the program recognizes 'EIN' (or one of its forms) it immediately looks at the next word, looks up its translation and examines the first letter of the translation. If the first letter is a vowel, the translation of the 'EIN-word' is 'AN' and if the first letter is a consonant, as it is here, the translation is 'A'.

(2) ICH HABE EINEN APFEL.
I HAVE AN APPLE.

Again this sentence requires, for the most part, straight translation, except for the German word 'EINEN' (the accusative masculaine form of 'EIN'), which can be translated as either 'A' or 'AN'. The program again examines the translation of the next word of the sentence. Because in this case the translation begins with a vowel, 'AN' is chosen as the proper translation.by the program.

(3) SPIELT DER KNABE BALL*
DOES THE BOY PLAY BALL*

This sentence is an example of one form of German question. Schematically, the German form is

VERB + SUBJECT + PREDICATE *

while the English form is

DOES + SUBJECT + VERB(-S) + PREDICATE *

This is the first example where the arrangement problem is met. The program handles this example the following way. While reading the input, the program looks at the end of the sentence for either a period or a question mark (*). If it encounters a question mark, the next step is to look at the first word of the sentence. In the question form of this example the first word will always be a verb. The program must then place a form of 'DO' as the first word of the translation. To determine which form (either 'DO' or 'DOES') the program looks at the verb ending. If the ending is a '-t', not preceded by an 's', the verb will almost always be third person singular, and the appropriate form would be 'DOES'. In all other cases the correct form would be 'DO'. It should be noted here that when 'DOES' is chosen, the translated verb will have the 's' of its ending deleted.

The next step is to rearrange the word order to insert the verb in the proper place. Here the subject marker (+) is exploited. The verb, which is in place 1 of the sentence array, is put in storage, and each word in the sentence up to the subject marker is moved up one place in the sentence array. When the location of the subject marker is reached, the verb is retrieved from storage and inserted into the empty place in the array. The look-up and resulting translation then proceeds normally.

Thus, in this sentence, following the discovery of the question mark, the program notes that the first word is a verb in the third person singular ('SPIELT') and inserts 'DOES' into the translation. It then notes that the subject marker follows 'KNABE', moves all the words up to and including 'KNABE' one step up in the sentence array, and inserts 'SPIELT'. When the translator reaches 'SPIELT', it notes that 'DOES' has been used and deletes the 's' from 'PLAYS' (the English translation of 'SPIELT'). The rest of the sentence is translated routinely.

(4) SCHWIMMEN WIR MIT IHR*
DO WE SWIM WITH HER*

This sentence is of the same type as (3) above. The verb ('SCHWIMMEN'), however, is first person plural, so 'DO' is used instead of 'DOES'. The rest of the translation proceeds as did that for (3), until the word 'IHR' is reached. Here another selection decision must be made. The pronoun 'IHR' has more than one meaning. In the nominative case 'IHR' means 'YOU' in the plural, while in the dative it means 'HER'. When 'IHR' is encountered in the sentence, its location is examined. If it is in the subject, it will only be translated as 'HER' if preceded by a dative preposition; otherwise it will be translated as 'YOU'. Similarly, if 'IHR' appears in the predicate it will be translated as 'HER' if preceded by a dative preposition in the predicate, or if the verb is not a form of 'TO BE'. 'YOU' is the translation only if there is no dative preposition in the predicate preceding it and if the verb is 'TO BE'..

(5) WO IST DAS KLEINE MAEDCHEN *
WHERE IS THE SMALL GIRL *

This question has as its first word 'WO' which is translated as 'WHERE'. The 'wh-words' are translated routinely and otherwise ignored in the formulation of English translations. Thus, following the translation of the 'wh-word', the rest of the question itself is treated as the question, with 'DO' or 'DOES' inserted as needed. The above sentence has a form of 'TO BE' as its verb and is translated routinely.

(6) IST DER HIMMEL BLAU *
IS THE SKY BLUE *

This sentence is very similar to (5) above, except there is no 'wh-word'. When the program recognizes the question mark, if it discovers the first word to be a form of 'SEIN' ('TO BE'), it translates the sentence word for word, noting only that there are no words in the predicate. This is what is done here.

(7) SIE LAUFEN IN DAS HAUS.
THEY RUN INTO THE HOUSE.

This sentence has two interesting examples of the selection problem. The sentence is translated with no rearranging of the words, but the proper translations of 'SIE' and 'IN' must be made.

'SIE' is a German pronoun that can either mean 'SHE' (in the nominative), 'THEY' (in the nominative plural), 'HER' (in the accusative) or 'THEM' (in the accusative plural). There is first no syntactic means of distinguishing between 'HER' and 'THEM', so the system, when faced with that choice always chooses 'HER'.. The other choices are made this way. Anytime 'SIE' follows an accusative preposition, it is translated as 'HER'; as it is whenever it is in the predicate of a sentence with a verb other than 'TO BE'. If it appears anywhere in a sentence that contains 'TO BE', or in the subject of any sentence, 'SIE' is translated as 'SHE' or 'THEY', depending on the verb-ending. If the ending is '-EN' or if the verb is plural form of 'SEIN', the choice will be 'THEY', otherwise it will be 'SHE'.

In this sentence, the program sees that 'SIE' appears in the subject of a sentence whose verb form ends with '-EN'. Thus, it translates 'SIE' as 'THEY'.

The next selection that must be made is in choosing the proper translation of 'IN'. 'IN' is a German preposition that takes either the dative or accusative cases. When it is in the dative, it connotes a stationary physical state, and is translated as 'IN'. When it takes the accusative, it implies movement and is translated as 'INTO'. To make this selection decision the program looks at the definite or indefinite article following 'IN'. If the article is dative, 'IN' is translated as 'IN', and if the article is accusative, 'IN' is translated as 'INTO'.

In the above sentence 'IN' is followed by the nominative-accusative neuter definite article 'DAS'. Consequently, the program correctly translates 'IN' as 'INTO'.

(8) DER KLEINE KNABE LAEUFT HINTER SIE.
THE SMALL BOY RUNS BEHIND HER.

This sentence translates routinely with no need for rearranging. There is, however, a selection decision to be made when the translator reaches 'SIE'. As in the discussion for (7), 'SIE' has several English alternatives. The key here is that 'SIE' follows a preposition that can be either dative or accusative. Because 'SIE' does follow it, it is, of necessity, accusative. Hence, the English equivalent to 'SIE' in this case is 'HER'.

(9) SIE WIRD SEHR GUT.
SHE BECOMES VERY GOOD.

This sentence is translated with a selection decision for the very first word 'SIE'. Here it appears in the subject (The subject marker shows that.) and the verb has no '-EN' ending. Hence, it is correctly translated as 'SHE', as discussed in (7) above.

The other word worth noting here is 'WIRD'. 'WIRD' can either mean 'BECOMES' as it does here, or it can be used as the auxiliary for the future tense. For that usage the infinitive form of the actual verb must appear at the end of the sentence. Thus, when 'WERDEN' or one of its forms is encountered, the program immediately searches the last word of the sentence for an infinitive verb form. This sentence has none, so 'WIRD' is translated as 'BECOMES'.

(10) ICH WERDE MIT IHM SCHWIMMEN.
I WILL SWIM WITH HIM.

This sentence is a simple example of a future tense sentence. Here the program has seen 'WERDE' (a form of 'WERDEN') and looked for the infinitive verb form at the end of the sentence. It finds 'SCHWIMMEN' and performs the needed transformation.

SUBJECT + WERDEN + PREDICATE + INF. VERB

in German becomes,

SUBJECT + WILL VERB + PREDICATE

in English.

The program handles this by placing 'WILL' with the verb following the translation of the subject. This location has been obtained by use of the subject marker. The program then totally eliminates from the sentence array the original infinitive verb. This is what was done for this sentence.

(11) DER HUND IST SCHNELL.
THE DOG IS → SCHNELL.

This is a simple sentence which shows what the program does when a word does not appear in the German lexicon. The above translation would appear with the message 'THE WORD SCHNELL DOES NOT APPEAR IN OUR LEXICON'.

(12) DAS MAEDCHEN SCHWIMMT IN DEM WASSER.
THE GIRL SWIMS IN THE WATER.

This is a simple straightforward sentence. The only selection decision has to deal with the word 'IN'. 'IN' is followed by the dative definite article 'DEM', and as was seen in (7) above, when 'IN' is in the dative it is translated as 'IN'.

(13) DER HUND DER FRAU LAEUFT.
THE DOG OF THE WOMAN RUNS.

This sentence shows a selection decision that has been performed throughout, but has not been discussed until now. That has to deal with the genetive construction. The definite articles 'DER' and 'DES' are the definite articles for the genetive case. The genetive case is used in German as the possessive construction. The way the program recognizes this is by looking at the two appropriate definite articles. If either 'DER' or 'DES' is preceded by a noun, the construction is taken to be possessive and the definite article is then translated as 'OF THE'. In the case of 'DES', which is used with masculine and neuter nouns, the noun usually adds an 's' or 'es'. These must be deleted when the dictionary is searched for a match in the look-up phase of the program.

In the above sentence the first occurrence of 'DER' has no noun in front of it and is thus translated simply as 'THE'. The second occurrence of 'DER', however, is preceded by a noun and is thus translated as 'OF THE'. The rest of the sentence is translated routinely.

(14) DAS BUCH DES VATERS IST GUT.
THE BOOK OF THE FATHER IS GOOD.

This sentence again uses the genetive construction. 'DES', preceded by a noun, is translated as 'OF THE'. The only other manipulation that must be performed on this sentence is the deletion of the 's' from 'VATERS' before the dictionary look-up for 'VATER'. The program does this by looking at the definite article 'DES', recognizing the genetive case, and deleting the trailing 's' from the following noun.

7. Application of the German-English Mechanical Translator toward Computer-Assisted Instruction

For the German-English translation program to be most effective the student should be able to interact with it. This is not possible to a great extent at Lehigh due to the severe taxing of the storage capabilities put on the CDC 6400 by the SNOBOL4 interpreter. Consequently, any application of the German-English translator to CAI at Lehigh would be limited. (of necessity)

Largely to the batch mode.

Even here, however, the system could be put to great use in aiding beginning students. The fact that the student would see the output thirty minutes later rather than immediately would not be a hindrance to the success of the program.

The postive aspects of CAI for beginning German at Lehigh are basically the same ones that typify CAI anywhere. Explaining translations takes much time for the instuctor, so any help the mechanical translator could give to the student would be of some aid to the instructor. The ability to practice translating would also be quite helpful to the beginning student, at a time when he needs all the drilling he can possibly get.

That the student must also be able to pick out the subject of many sentences for the program to be effective, is an additional drill for the student. Without the basic understanding of the structure of the sentence, the student will place the subject marker improperly and will receive nonsense translations in many cases.

The success of the program as a teaching aid will be greatly increased by the ability of the program to compare its translation of its sentence with one the student types in himself. Once the machine has been programmed to recognize common errors, it can print an appropriate error message.

The CAI experiment at SUNY at Stony Brook that was mentioned earlier points to many benefits and possible causes for the success of it and a potential system using the German-English translating program developed here. One reason for the success seems to be, quite simply, a "fun factor". It is fun to learn by playing with a machine. This may be the reason for the low withdrawal rate in the SUNY study, especially the first semester when the experience is still novel. Thus a benefit is simply getting the student to work more on the elementary concepts of the language. Another benefit of the system is one of bolstering the morale of the beginning student who is often overwhelmed by learning a new language. There is the feeling, "if this machine can do it, so can I." This is quite helpful for the beginning student.

Thus, the application of the German-English translator as a teaching aid for beginning students has great possibilities. The only restrictions on it would be imposed by cost and the difficulty of translation.

Those wishing to examine the translation program may do so at the Division of Information Science in the Philosophy Building.

References

Alt, F. L. and M. Rubinoff eds.
1871 Advances in Computers, Vol. 11, (New York, Academic Press)

Atkinson, R. C. and H. A. Wilson eds.
1972 Computer-Assisted Instruction, (New York, Academic Press)

Booth, A. D. ed.
1867 Machine Translation, (New York, John Wiley)

Booth, K.
1967 "Machine-Aided Translation with a Post-Editor", in Machine Translation, A. D. Booth ed., (New York, John Wiley), pp. 51-76.

Dostert, L. E.
1955 "The Georgetown-I.B.M. Experiment", in Machine Translation of Languages, W. N. Locke and A. D. Booth eds., (Cambridge, M. I. T. Press), pp. 124-135.

Garvin, P. A.
1956 "Some Linguistic Problems in Machine Translation: An Early View Still Held", in On Machine Translation, by P. A. Garvin, (The Hague, Mouton), pp. 65-75.
1958 "Machine Translation: A Report to the VIII International Congress on Linguistics", in On Machine Translation, (The Hague, Mouton), pp. 76-82.
1968 "Machine Translation Today: The Fulcrum Approach and Heuristics", in On Machine Translation, (The Hague, Mouton), pp. 90-116.
1972 On Machine Translation, (The Hague, Mouton)

Griswold, R. E., J. F. Poage and I. P. Polansky
1971 SNOBOL4 Programming Language, (Englewood Cliffs, N. J., Prentice-Hall)

Josselson, H. H.
1971 "Automatic Translation of Languages Since 1960: A Linguist's View", in Advances in Computers, Vol. 11, F. L. Alt and M. Rubinoff eds., (New York, Academic Press), pp. 2-59.

Karlowiak, A. E. and R. M. Huey eds.
1971 Information, Computers, Machines and Man, (New York, John Wiley)

Locke, W. N. and A. D. Booth eds.
1855 Machine Translation of Languages, (Cambridge, M. I. T. Press)

Lowe, T. L., J. J. Mulligan and A. H. Wegener
1971 Modern College German, (New York, Harper and Row)

Masterman, M.
1967 "Mechanical Pidgin Translation", in Machine Translation, A. D. Booth ed., (New York, John Wiley), pp. 195-228.

Meadow, C. T.
1970 Man-Machine Communication, (New York, John Wiley)

Morrison, H. W. and E. N. Adams
1972 "Pilot Study of a CAI Laboratory in German", in Computer-Assisted Instruction, R. C. Atkinson and H. A. Wilson eds., (New York, Academic Press), pp. 186-201.

Oettinger, A. G.
1855 "The Design of an Automatic Russian-English Technical Dictionary", in Machine Translation of Languages, W. N. Locke and A. D. Booth eds., Cambridge, M. I. T. Press), pp. 47-65.

Plath, W. J.
1967 "Multiple-Path Analys and Automatic Translation", in Machine Translation, A. D. Booth ed., (New York, John Wiley), pp. 267-316.

Reifler, E.
1855 "The Mechanical Determination of Meaning", in Machine Translation of Languages, W. N. Locke and A. D. Booth eds., (Cambridge, M. I. T. Press), pp. 136-164.

Singh, J.
1966 Great Ideas in Information Theory, Language and Cybernetics, (New York, Dover Publications)

Weaver, W.
1955 "Translation", in Machine Translation of Languages, W. N. Locke and A. D. Booth eds., (Cambridge, M. I. T. Press), pp. 15-23.

Appendix I: Vocabularies

GERMAN	ENGLISH
DER	THE
DIE	"
DAS	"
DEN	"
DEM	"
DES	"
EIN	A, AN
KNABE	BOY
MANN	MAN
FRAU	WOMAN
MAEDCHEN	GIRL
APFEL	APPLE
HAUS	HOUSE
HUND	DOG
BERG	MOUNTAIN
BALL	BALL
MUTTER	MOTHER
VATER	FATHER
SCHWESTER	SISTER
BRUDER	BROTHER
BUCH	BOOK
PERSONN	PERSON
LEUTE	PEOPLE
WASSER	WATER
HIMMEL	SKY
ER	HE
ES	IT
ICH	I
DU	YOU
WIR	WE
SIE	SHE, THEY, HER, THEM
IHR	YOU, HER
MICH	ME
DICH	YOU
EUCH	"
IHN	HIM
IHNEN	THEM
UNS	US
MIR	ME
DIR	YOU
IHM	HIM
AUS	OUT
BEI	BY

GERMAN	ENGLISH
MIT	WITH
NACH	AFTER
VON	FROM
ZU	TO
IN	IN, INTO
HINTER	BEHIND
ZWISCHEN	BETWEEN
GUT	GOOD
ROT	RED
BLAU	BLUE
SCHLECHT	BAD
KLEIN	SMALL
GROSS	LARGE
SCHOEN	PRETTY
SEHR	VERY
JA	YES
NEIN	NO
NICHT	NOT
UND	AND
ODER	OR
ABER	BUT
SONDERN	BUT
AUCH	ALSO
WAS	WHAT
WARUM	WHY
WO	WHERE
WOHIN	WHERE
WIE	HOW
WER	WHO
WEM	WHOM
WEN	WHOM
SEIN	TO BE
WERDEN	TO BECOME, WILL
HABEN	TO HAVE
SPIELEN	TO PLAY
LAUFEN	TO RUN
SCHWIMMEN	TO SWIM

Appendix II: Sample Sentences, German Input and German and English Outputs

ICH + HABE EINE SCHWESTER .
ICH HABE EINE SCHWESTER .

THE TRANSLATION OF THIS SENTENCE IS:
I HAVE A SISTER .

ICH + HABE EINEN APFEL .
ICH HABE EINEN APFEL .

THE TRANSLATION OF THIS SENTENCE IS:
I HAVE AN APPLE .

SPIELT DER KNABE + BALL *
SPIELT DER KNABE BALL *

THE TRANSLATION OF THIS SENTENCE IS:
DOES THE BOY PLAY BALL *

SCHWIMMEN WIR + MIT IHR *
SCHWIMMEN WIR MIT IHR *

THE TRANSLATION OF THIS SENTENCE IS:
DO WE SWIM WITH HER *

WO IST DAS KLEINE MAEDCHEN + *
WO IST DAS KLEINE MAEDCHEN *

THE TRANSLATION OF THIS SENTENCE IS:
WHERE IS THE SMALL GIRL *

IST DER HIMMEL + BLAU *
IST DER HIMMEL BLAU *

THE TRANSLATION OF THIS SENTENCE IS:
IS THE SKY BLUE *

SIE + LAUFEN IN DAS HAUS .
SIE LAUFEN IN DAS HAUS .

THE TRANSLATION OF THIS SENTENCE IS:
THEY RUN INTO THE HOUSE .

DER KLEINE KNABE + LAEUFT HINTER SIE .
DER KLEINE KNABE LAEUFT HINTER SIE .

THE TRANSLATION OF THIS SENTENCE IS:
THE SMALL BOY RUNS BEHIND HER .

SIE + WIRD SEHR GUT .
SIE WIRD SEHR GUT .

THE TRANSLATION OF THIS SENTENCE IS:
SHE BECOMES VERY GOOD .

ICH + WERDE MIT IHM SCHWIMMEN .
ICH WERDE MIT IHM SCHWIMMEN .

THE TRANSLATION OF THIS SENTENCE IS:
I WILL SWIM WITH HIM .

DER HUND + IST SCHNELL .
DER HUND IST SCHNELL .

THE TRANSLATION OF THIS SENTENCE IS:
THE DOG IS — SCHNELL .
THE WORD SCHNELL DOES NOT APPEAR IN OUR LEXICON

DAS MAEDCHEN + SCHWIMMT IN DEM WASSER .
DAS MAEDCHEN SCHWIMMT IN DEM WASSER .

THE TRANSLATION OF THIS SENTENCE IS:
THE GIRL SWIMS IN THE WATER .

DER HUND DER FRAU + LAEUFT .
DER HUND DER FRAU LAEUFT .

THE TRANSLATION OF THIS SENTENCE IS:
THE DOG OF THE WOMAN RUNS .

DAS BUCH DES VATERS + IST GUT .
DAS BUCH DES VATERS IST GUT .

THE TRANSLATION OF THIS SENTENCE IS:
THE BOOK OF THE FATHER IS GOOD .

Appendix III: User's Manual

The following are some rules for the use of the German-English mechanical translator.

1. Type one sentence per card.
2. Put at least one space between words and between a word and a punctuation mark.
3. Use a plus sign (+) as the subject marker. That is, place it after the last word of the subject.
4. Use an asterisk (*) as a question mark.
5. Use the words in the vocabulary.
6. Type 'XXX' on the last card of your input.
7. Place your input cards following the 'END' card of the SNOBOL deck, and in front of the 'EOF' card.

Appendix IV: System Manual

The following are the procedures one should follow if adding to the German-English translation program.

1. Add vocabulary by entering a card that reads as follows:
   WORD ' ZZZZ ' = 'YYYY '
   where 'XXXX' is the German word to be added and 'YYYY' is its English translation.

2. Normal entries should branch to PT. Any entry that requires special treatment should branch to a place at which the required manipulations are performed.

3. Be sure to add nouns to the appropriate noun list also.

4. All verbs should be added to the verb lists.

5. If adding any new capabilities, be sure that the already written programming is revised.

Vita

Mark Alan Bendas was born in Perth Amboy, New Jersey on August 2, 1950 to Charles and Mary Bendas. He was educated in the Perth Amboy public school system and graduated Perth Amboy High School in 1968.

He attended Lehigh University majoring in mathematics. While a student, he was named to the Dean's List, received Sophomore Honors and was elected to Omicron Delta Kappa national leadership honorary. He received his B. A. with honors in mathematics in 1972.

He returned to Lehigh in the Fall of 1972 to study for his M. S. in Information Science. While a graduate student he has served as a teaching assistant for Logic.